



Le pipeline de mes rêves...

E.Bertin, (IAP & Obs. de Paris/LERMA)



Conditions aux limites

- Pouvoir gérer plusieurs types d'instruments/survey sans avoir à modifier le code du pipeline
 - Grands surveys
 - Données P.I.
- Robustesse vis-à-vis des pannes
- Optimiser au maximum l'usage du pool de machines disponibles (à la fois CPU et I/O)
- Maximiser l'indépendance envers les outils de traitement des données
- Sécurité "raisonnable": comptes indépendants pour les pipelines et programmes P.I.
- Code distribuable?
- 2 écueils:
 - Pipeline *a priori* simple mais *a posteriori* trop ciblé (pipeline EIS)
 - Pipeline *a priori* générique mais *a posteriori* trop compliqué (pipeline AstroWISE)



Pouvoir gérer plusieurs types d'instruments/survey sans avoir à modifier le code du pipeline

- Si SPICA ne le fait pas, d'autres le feront ("pipeline" de Bonn)
- Indépendance entre le contenu de la base de données pipeline et celui des headers FITS
 - Formulaire de mapping entre mots-clé FITS et un jeu prédéfini de champs de la base de données.
 - Le résultat fait partie du jeu de préférences disponible pour un processing
 - Possibilité de remplacer/ajouter des contenus particuliers de mots-clés
 - Possibilité d'utiliser des fichiers du genre .head en entrée?
 - Possibilité de mélanger plusieurs relevés au sein du même processing
- Gestion des poses uniquement au niveau MEF
 - Plus de pb désormais avec les outils de bas niveau
- Sélection intuitive des champs à calibrer/stacker
 - SkyWatcher pour la sélection en alpha,delta
 - Concept de panier
 - Outils de réplcation
 - Calcul "temps réel" de statistiques: surface couverte, temps de pose équivalent,etc.



Robustesse vis-à-vis des pannes et autres problèmes de processing

- Sommes de contrôle MD5 pour toutes les opérations de transfert de (gros) fichiers entre machines
 - Plus rassurant pour les calculs distribués de type grid-computing
- Timeouts ajustables
- Contrôle automatique de l'intégrité des fichiers produits par les outils de processing
- Logs et historique des crashes disponibles pour l'utilisateur sur une page spéciale
- Base temporaire de stockage des informations intermédiaires (crash-sensitive)
- En cas de panne le processing continue
 - sur d'autres noeuds en cas de panne de machine/disque
 - Sur d'autres stacks en cas de panne logicielle spécifique aux données
- Possibilité de réinitialiser une étape de processing (attente de validation)
 - Autoriser plusieurs versions des résultats?



Optimiser au maximum l'usage du pool de machines disponibles (à la fois CPU et I/O)

- Gérer indépendamment les calibrations astrométriques et photométriques d'une part, et le rééchantillonnage/stacking d'autre part
 - Possibilité de lancer un SCAMP sans nécessairement créer un stack.
- Antagonisme entre l'efficacité "ponctuelle" du processing et celle de la répartition des charges sur le long terme (plusieurs semaines)
 - Usage 24h/24
 - Impact des différentes possibilités de pannes
 - Impact relatif des I/O et de la répartition des tâches
- Nouvelle donne: webservices
 - Les outils du pipeline peuvent être mis à contribution par des utilisateurs extérieurs
 - Gestion de priorité



Maximiser l'indépendance envers les outils de traitement des données

- Possibilité de choisir une version spécifique de chaque outil (SCAMP, SWarp, etc.) au sein du panier de processing
 - Utile pour les tests
- Pages de diagnostic XML+XSLT générées par les outils de processing eux-même, pouvant évoluer avec le no de version.
- Usage sous forme d'un web-service
 - Gestion des versions?



Sécurité et gestion des comptes

- Comptes indépendants pour les pipelines et programmes P.I.
 - Liste d'e-mails associée à chaque compte
 - Possibilité d'uploader/downloader des données (P.I. par exemple)
 - Affichage explicite de ce qui est disponible sur les disques et ce qui ne l'est plus
 - Pas de lien cassé



SPICA distribuable

- Sous forme de code ou de web-service.
- Un tel pipeline apparait faisable. Il sera donc probablement fait par d'autres si cela n'est pas fait à TERAPIX (exemple: Bonn)
- Etendre la base installée
 - Améliorer la portabilité
 - Fédération des outils et web-services TERAPIX
 - Visibilité TERAPIX



terapix.iap.fr

